Social Network Analysis of Gene Expression Data

Jung Hun Ohn, M.D., Jihoon Kim, B.S., Ju Han Kim, M.D., Ph.D.*

SNUBI: Seoul National University Biomedical Informatics, SNU School of Medicine, Seoul 110-799, Korea

Abstract

To investigate the structure of genomic interaction network, social affiliation network analysis was performed for the yeast gene expression compendium dataset of hundreds of systematic perturbations. Network density and centrality indices of genes and groups of genes revealed the core-peripheral and the significant intermediary players that may be critical for the control of the biological system.

Introduction

Cellular processes are mainly described with mechanical understandings. Cellular processes are regarded as conveyer belts many workers (or proteins) work on to give products essential for the survival of a large factory (or a cell). Individual biological building blocks such as genes or proteins, however, may not possess the explicit understanding of the purpose or intention of what they perform in the context of cellular process. The notion of cellular process as an emergent property of these purposeless interactions may in fact be a better description of life. We have applied the wisdom of social analogy for the analysis of gene expression data. By means of social network analysis, we could infer the property of individual genes through their social groups and relationships: we know a gene by the company it keeps.

Methods and Results

By means of one-mode and two-mode analysis [1] with the analogy of genes as actors and experimental conditions as events represented in an affiliation matrix, we identified the social 'stars' and the network structures from the Rosetta Compendium dataset [2] with 300 diverse mutations and chemical treatments in *S. cerevisiae*. Bipartite matrix and bipartite graph were constructed. We measured the rates of participation, size of events, and the centrality indices such as node degree, closeness, and betweenness centralities. Group centralization measures, clustering structure and core-periphery structure were determined in company with MIPS functional classification [3] to explore the genome-wide interaction structure. The 'social-star' genes participating in a large number of events seemed to belong to the essential stress-responsive categories: stress response, amino acid biosynthesis, C-compound and carbohydrate biosynthesis, small molecule transport, and osmoregulation. Ribosomal biogenesis, lipid, fatty-acid and isoprenoid biogenesis, transport, transcriptional control, cell cycle, DNA synthesis and replication, budding and pheromone response showed the largest event sizes. For the purpose of illustration, we have explored the core-periphery structures of the energy-related processes and TCA cycle. Interestingly, genes participating in fatty acid oxidation and energy transport are mostly placed in periphery, whereas glucose-metabolism related processes contain core genes in energy process and ATP generating processes occupy intermediate position. Glyoxylate cycle gene group has the highest degree and closeness centrality and exhibited the most 'star' shaped graph. In contrast, respiration process has the smallest degree, closeness, and betweenness centrality indices and showed the least 'star' like structure. Among the 20 actors of TCA cycle from MIPS, the core group contains the well-known succinate dehydrogenase complex, isocitrate dehydrogenase, fumarase, aconitase, and citrate syntase while genes assigned to the periphery group have hitherto unspecified role in TCA cycle.

Discussion

Large-scale gene expression profile was investigated in the context of the social network analysis where genes are regarded as actors, conditions as events, and the network topology as variety of centrality and relatedness indices. The analysis demonstrated some important features such as core-peripheral players and significant intermediary actors that may be critical for the control of the system and useful for the development of valuable therapeutic substances.

References

[1] Borgatti SP, Everett MG. Network analysis of 2-mode data. Social Networks 1997;19:243-269

[2] Hughes TR, Marton MJ, Jones AR et al. Functional discovery via a compendium of expression profiles. Cell 2000 Jul 7;102(1):109-26.

[3] http://mips.gsf.de

* This study was supported by a grant from Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (01-PJ10-PG6-01GM01-0004). Letter of correspondence to juhan@snu.ac.kr